Statistical Exploratory Analysis of Genetic Algorithms

Andrew Czarn, Cara MacNish, Kaipillil Vijayan, Berwin Turlach, and Ritu Gupta

***Abstract*—Genetic algorithms have been extensively used and studied in computer science, yet there is no generally accepted methodology for exploring which parameters significantly affect performance, whether there is any interaction between parameters, and how performance varies with respect to changes in parameters.**

This paper presents a rigorous yet practical statistical methodology for the exploratory study of genetic and other adaptive algorithms. This methodology addresses the issues of experimental design, blocking, power calculations, and response curve analysis. It details how statistical analysis may assist the investigator along the exploratory pathway. As a demonstration of our methodology, we describe case studies using four well-known test functions.

We find that the effect upon performance of crossover is predominantly linear, while the effect of mutation is predominantly quadratic. Higher order effects are noted but contribute less to overall behavior. In the case of crossover, both positive and negative gradients are found suggesting the use of a maximum crossover rate for some problems and its exclusion for others. For mutation, optimal rates appear higher compared with earlier recommendations in the literature, while supporting more recent work. The significance of interaction and the best values for crossover and mutation are problem specific.

***Index Terms*—Adaptive algorithms, experimental design, genetic algorithms (GAs), methodology, statistical analysis.**

I. Introduction

ADAPTIVE algorithms such as genetic algorithms (GAs) [1] work by iteratively adapting members of a population of potential solutions. The individuals interact either through the adaptation operators themselves, or through competitive selection mechanisms for determining subsequent generations. If the adaptation strategy is successful, the population (or part thereof) will converge on an optimal (or at least "good") solution.

While the mechanics of each individual adaptation are quite straightforward, the way individual changes affect the success of the population as a whole is more difficult to determine. This is also true of the many parameters that are used to fine tune, or improve the success of adaptive algorithms. Examples include population size, mutation and crossover rates, elite group sizes, acceleration constants, step sizes, and so on. Values for these parameters are most commonly set through a process of trial and error, or based on recommendations from related problems in the literature, rather than through statistically sound analysis of their effects on algorithm performance.

In this paper, we propose a rigorous yet practical statistical methodology for assessing the impact of parameter settings. The methodology addresses issues of experimental design, blocking, power calculation, and response curve analysis. We demonstrate the approach with a case study applying GAs to benchmark problems from De Jong's [2] and Schaffer's [3] test suites.

In Section II, we provide some background to the problem of analyzing GA performance. This is followed in Section III by a discussion of nonstatistical exploratory work in this area. Section IV examines work which has used a statistical construct, recognizing the appropriateness of statistical analysis to this problem. However, a number of limitations are found. In Section V, we discuss a range of factors that must be considered in developing a suitable methodology and outline our approach. The results of applying this methodology to the GA in our case study are reported in Section VI. This includes some unexpected outcomes, particularly on the use of crossover. A discussion in Section VII concludes the paper.

II. Background

A GA works by encoding potential solutions to a problem as a series of bits or *genes* on a bit-string or *chromosome*. The mechanics of a GA are straightforward: in its simplest form new solutions are generated using *crossover*, where genes are crossed over between pairs of chromosomes, and *mutation*, where the binary value of a gene is inverted.

In contrast, the *way* in which a GA population converges on solutions has been more complex to describe [1]. Holland put forward the idea of *schemata* [4]: similarity templates describing a subset of strings with similarities at certain positions [5]. When the chromosome possesses these schemata its fitness improves. Operators such as crossover and mutation work by altering chromosomes to contain more good schemata. Goldberg elaborated by conceptualizing *building blocks* (highly fit, short-defining-length schemata) and *implicit parallelism* [5]. However, the increase in sophistication and differences in implementations of GAs, such as quantum-inspired GAs [6] and the use of transposition [7], has made it increasingly difficult to propose newer models of convergence.

In addition, previously accepted aspects of GAs are being debated. For example, while it has been traditionally maintained that crossover is a necessary inclusion, the conjecture of *naive evolution* (using selection and mutation only) places

Manuscript received September 8, 2003; revised April 30, 2004.

A. Czarn and C. MacNish are with the School of Computer Science and Software Engineering, University of Western Australia, Crawley 6009, Western Australia (e-mail: aczarn@csse.uwa.edu.au; cara@csse.uwa.edu.au).

K. Vijayan and B. Turlach are with the School of Mathematics and Statistics, University of Western Australia, Crawley 6009, Western Australia (e-mail: vijayan@maths.uwa.edu.au; berwin@maths.uwa.edu.au).

R. Gupta is with the Department of Mathematics and Statistics, Curtin University of Technology, Bentley 6102, Western Australia (e-mail: ritu@maths.curtin.edu.au).

Digital Object Identifier 10.1109/TEVC.2004.831262

this in question [8], [9]. Such debates have been fuelled by the fact that little research has been done on how to decide whether a parameter significantly affects performance and how performance varies with respect to changes in parameters. There is currently no generally accepted methodology for exploring a GA in order to address these issues.

The difficulty in developing such a methodology is illustrated by problems encountered in both working from theoretical models and real-world data. In the first instance, trying to formally describe GAs has been attempted using various mathematical approaches such as Markov chains [10], [11]. These approaches have been limited by the complexity of the calculations. Moreover, the assumptions made in much of the theoretical work may simply not be applicable nor attainable in practice, such as assuming an infinite population size when considering the processing of schemata. There has, therefore, been a realization that research involving real-world data will be necessary in order to provide guidelines that may come to be generally accepted by GA practitioners.

Initial empirical work of this kind was carried out by De Jong [2] whose experiments resulted in a set of recommendations that came to represent early guidelines [8]. Later recommendations by Grefensette [12] using a *meta-level* genetic algorithm (meta-GA) produced results which did not wholly agree with De Jong. The meta-GA approach is limited in that independent runs of the meta-GA can result in different best values. Furthermore, it does not provide any information as to whether any interaction occurs nor the trend of the performance behavior over the range of values studied.

A limited number of studies have made use of statistical analysis, recognizing the ability of statistics to address many of these issues. However, as discussed in Section IV, these studies have been limited by failing to fully address important issues such as blocking for seed, calculating power, and thorough response curve analysis. Thus, results and recommendations from these studies, though obtained from real-practical experience, are still subject to debate.

We describe a statistical methodology for such exploratory work with real-world data. This methodology is rigorous yet practical with general principles that can be applied to the practical analysis of other kinds of adaptive algorithms.

In the next sections, we look more closely at the various studies in this area. In doing so, we note the inconsistency of the results and the limitations of the methodologies. We then define our experimental setup and describe our statistical methodology.

III. Nonstatistical Exploratory Analysis

As stated above, there is currently no generally accepted methodology for analyzing the relationship between parameters and performance of a GA. Attempting to mathematically describe GAs is complex and has not resulted in practical guidelines. This has given rise to various studies which attempt to provide such data. However, both the methodologies and results have varied.

Early work was provided by De Jong who altered the values of parameters such as population size, crossover rate and mutation rate in order to assess the effect on performance. This was defined in terms of *online performance*, the average performance of all chromosomes tested during the search, and *offline performance*, the current best chromosome value for each iteration [8]. Five test problems of increasing difficulty were used which became known as the De Jong suite [2]. Table I lists De Jong's recommendations for optimal performance for the parameters listed.

TABLE I
Recommendations for Basic Parameter Settings

De Jong	Population size	50-100
	Crossover rate	0.60
	Mutation rate	0.001
Grefensette	Population size	30 (online)
	Population size	80 (offline)
	Crossover rate	0.95 (online)
	Crossover rate	0.45 (offline)
	Mutation rate	0.01 (online)
	Mutation rate	0.01 (offline)
Freisleben and Härtfelder	Population size	100 (maximal)
	Crossover rate	0.49
	Mutation rate	0.8-0.93

TABLE II
Recommendations for Basic Parameter Settings Using Statistics

Schaffer et al	Population size	20-30 (online)
	Crossover rate	0.75-0.95 (online)
	Mutation rate	0.005-0.01 (online)
Petrovski, Wilson and McCall	Crossover rate using Ψ	0.6146
	Mutation rate using Ψ	0.1981
	Crossover rate using $\log(\Psi)$	0.7600
	Mutation rate using $\log(\Psi)$	0.1069

At this stage, there was little evidence to dispel the idea that such data could serve as generic guidelines for different problem domains. Hence, these data came to represent guidelines for GA practitioners. Subsequent work, however, was not consistent with these recommendations.

This is illustrated in the results of Grefensette who pioneered the use of *meta-level* genetic algorithms (meta-GAs) [12] for finding optimal values for parameters. His results for the De Jong suite are shown in Table I. Other studies using the meta-GA approach also produced differing results, as seen in the work by Freisleben and Härtfelder [13] in the domain of neural network weights optimization (see Table I).

IV. Statistical Exploratory Analysis

As the previous studies did not clarify the relationship between parameters and performance statistical analysis has been used for this purpose. For example, Schaffer *et al.* [8] conducted a factorial design study using the analysis of variance (ANOVA). This study used the De Jong suite plus an additional five problems. The recommendations for best online performance from this study are shown in Table II. Close examination of the best online pools suggested a relative insensitivity to crossover which in turn suggested that naive evolution may be a powerful search algorithm in its own right when using bit string

encoding [8], [9]. Work by Yao *et al.* suggests that this may be also true when using real values [14]. These data challenge the traditional assumption that the crossover operator is a necessary inclusion in a GA [3].

Statistics was also used by Petrovski *et al.* [15] who carried out fractional factorial experiments in the domain of anti-cancer chemotherapy. These were combined with linear regression in order to pinpoint which parameters were significant and estimate their best values. The outcome measure Ψ was the number of generations required in order to reach the feasible region in the solution space. The results are shown in Table II.

In overview, it is clear from both the nonstatistical and statistical approaches that results have varied, notably for mutation where the more recent studies, including those using statistics, suggest higher rates. This may indicate a more complex effect for this parameter or alternatively that best values are problem specific. Moreover, the influence of differing problem domains must also be considered [16].

Importantly, however, the variation seen in these studies may also be a result of the differing methodologies that have been employed and, therefore, suggests the need to develop a generally accepted methodology for carrying out such exploratory work. While statistics is promising for this purpose, a number of limitations need to be addressed.

First, little attention has been given to blocking for seed as a source of variation or noise. As pointed out by Davis [17], finding good settings for parameters can be difficult due to the fact that the same parameter settings on the same problems can lead to different results. In practice, these differences can be traced to different pseudorandom number generator seeds in the initialization of populations and in the implementation of selection, crossover and mutation. Blocking for seed by grouping experimental units into homogenous blocks, so that each run of the GA for differing levels of crossover and mutation occurs with the same seeds, limits the cause of variation within blocks to the parameters under study. In this way, variation or noise is reduced and comparisons are sharpened [18].

Adding to this, issues dealing with the calculation of power and sample size have also largely been ignored. This has meant that it is uncertain whether the studies carried out have had adequate power and, thus, sample size to detect differences that could be considered noteworthy. Sample sizes which are too small will generally fail to result in statistical significance. This is particularly important if blocking is not carried out since the data-set is akin to a completely randomized design. In such a design, effects may not be detected due to the extent of background noise in the data-set produced by seed. Thus, a much larger sample size is required to detect effects of interest.

A detailed analysis of response curves has also been limited. It is important to undertake such an analysis as it allows one to study the behavior of the parameter over the range of values implemented. Such data are useful in the optimization process. For example, knowing that a parameter has a linear relationship to performance may suggest that either the value for the parameter is set as high as possible or that the parameter is excluded.

In Section V, we define our experimental setup and describe our statistical methodology.

V. Methods

Before describing our methodology, we briefly introduce the test functions and the algorithm used to illustrate our approach.

A. Choice of Standard Test Functions

It was important to select test functions which are well known. Initially, the first three problems from the De Jong [2] suite were tackled which are relatively easy for a GA to solve. This provided a useful set of problems, widely referenced in the literature, on which to demonstrate the initial applicability of our methodology. These were *F1* known as the SPHERE, *F3* known as the STEP function, and *F2* known as ROSENBROCK'S SADDLE.

We then proceeded to a more difficult problem and so chose the well known Schaffer's *F6* [3]. These were all implemented as minimization problems and are displayed in (1)–(4), respectively

$$f_1(\mathbf{x}) = \Sigma_{i=1}^{3} x_i^2, \quad -5.12 \le x_i \le 5.12 \tag{1}$$

$$f_3(\mathbf{x}) = \Sigma_{i=1}^{5} \lfloor x_i \rfloor, \quad -5.12 \le x_i \le 5.12 \tag{2}$$

$$f_2(\mathbf{x}) = 100\left(x_2 - x_1^2\right)^2 + (1 - x_1)^2, \quad -2.048 \le x_i \le 2.048 \tag{3}$$

$$f_6(\mathbf{x}) = 0.5 + \frac{\left(\sin\sqrt{x_1^2 + x_2^2}\right)^2 - 0.5}{\left(1.0 + 0.001\left(x_1^2 + x_2^2\right)\right)^2}, \quad -100.0 \le x_i \le 100.0. \tag{4}$$

B. Implementation of the GA

We implemented a GA as detailed in Table III. The implementation of the GA was deliberately simple so that a clear and concise demonstration of the proposed methodology and results could be made. In this regard, parameters such as the population size and bits per variable were not varied but kept at the values shown in Table III and only crossover and mutation were investigated in the present research. The same methodology can be straightforwardly applied to the many other parameters suggested in the literature.

C. Experimental Design and Statistical Test

In order to decide upon the most appropriate type of experimental design and statistical test, it was necessary to address several items:

1) blocking for variation or noise due to seed;
2) choice of an appropriate statistical test;
3) statistical testing of individual parameters and their interactions;
4) response curve analysis—this should allow for an estimate to be made of the best value for individual parameters with confidence intervals;
5) calculation of power;
6) a methodology that is rigorous yet practical enough to be undertaken with common statistical packages and available desktop computing power;
7) statistical principles that can be generically applied to other adaptive algorithms.

These are discussed in turn.

TABLE III
DETAILS OF THE GENETIC ALGORITHM

Variable representation	Bit string
Bits per variable	22
Genes	Binary value 1 or 0
Population size	50 chromosomes
Chromosome coding	Gray coding
Selection	Probabilistic selection [1]
Experimental unit	Blocks containing independent runs of the genetic algorithm for different crossover and mutation rates with the same seeds
Crossover	Single point (randomly selected) per variable
Mutation	Randomly generated bit replacement [2]
Performance measure	Final epoch ie epoch at which fitness of best chromosome $\leq 10^{-10}$ of maximum fitness for *F1*, *F2* and *F3* and epoch at which fitness of best chromosome $\leq 10^{-6}$ of maximum fitness for *F6*

[1]Probabilistic selection used here is the random selection of parents with the probability of selection being directly proportional to the fitness of a chromosome.

[2]Mutation is implemented as described by Davis [3]. That is, if the probability test is passed the binary bit is replaced by another binary bit that is randomly generated. Fifty per cent of the time the new bit will be the same as the old bit. The *bit-flipping mutation rate* is therefore half of the implemented mutation rate.

TABLE IV
CREATING A DATA-FILE FROM REPLICATES OF BLOCKS

Block	Parameter A	Parameter B	Observations
Seed/s for block-replicate 1	*i levels*	*j levels*	ij
Seed/s for block-replicate 2	*i levels*	*j levels*	ij
Seed/s for block-replicate 3	*i levels*	*j levels*	ij
$\vdots$	$\vdots$	$\vdots$	$\vdots$
Seed/s for block-replicate n	*i levels*	*j levels*	ij
Total observations $= ijn$ *where* $ij \geq 2$			

1) *Blocking*. The variation seen in GA runs is due to the differences in the starting population and the probabilistic implementation of mutation and crossover. This is in turn *directly* dependent on seed: the value used to generate the pseudorandom sequences. In usual implementations of a GA, the effect of seed is not regulated and so the experimental design may be conceived as being entirely randomized. In order to demonstrate statistically significant effects, a very large data-set is required in order to detect effects over and above variation or noise due to seed.

   To address this issue, it was necessary to control for the effect of seed via the implementation of a *randomized complete block* design. In such a design every combination of levels of parameters appears the same number of times in the same block and in the present study the blocks are defined through seeds. For example, if there are i levels of parameter A and j levels of parameter B, then each block contains all ij combinations.

   Seed is blocked by ensuring that the seeds used to implement items such as initialization of the starting population of chromosomes, selection, crossover, and mutation are identical within each block. An increase in sample size occurs by *replicating* blocks identical except for the seeds. Replicates of this type are necessary to assess whether the effects of parameters are significantly different from variation due to changes in seed. This is illustrated in Table IV.

2) *ANOVA*. In order to compare performances for two or more parameters using a randomized complete block design, we use the statistical test for the equality of means known as the analysis of variance (ANOVA). In ANOVA, the null hypothesis is that the means for different levels of a parameter are equal. The alternative hypothesis is that the means for levels of a parameter are not equal and, thus, we conclude that the parameter has an effect upon the response variable.

   ANOVA is so called as it essentially splits the total variation in the observations into variation contributed by the parameters (crossover and mutation), their interaction, block, and error. Error is conceptualized in terms of *residuals*, which are simply the individual deviations of the observations from the expected values based on the assumption that there is no effect.

   Testing to ascertain if a parameter such as crossover or mutation has a statistically significant effect is a straightforward process. First, we divide the variation contributed by the parameter adjusted by the number of levels of the parameter by the variation contributed by error adjusted by the number of levels of the parameters and the observations. This results in a ratio which is called an F value. Second, we determine the probability that we would observe an F value as large as we did under the null hypothesis. This is the p-value associated with the F value or simply $\Pr(F)$.

   If the p-value is equal to or less than a chosen level of significance (see Section V-D) this is taken to suggest that the parameter has an effect upon the response variable. A typical output from ANOVA is shown in Table VI. Here, it can be seen that crossover, mutation, and block would be considered to have an effect at a 1% level of significance.

   In ANOVA, the values for $\Pr(F)$ (p-values) are only (exactly) valid if the responses are normally distributed. Although even moderate departures from normality do not necessarily imply a serious violation of the assumptions on which ANOVA is based [19], particularly for large sample sizes, it is standard procedure to use methods such as plotting a histogram of the residuals or constructing a normal probability plot of the residuals to verify normality of the sampling populations. In the present research analysis of the residuals did not provide any evidence suggesting that the assumptions on which ANOVA calculations are made were compromised.

3) *Testing individual parameters and interaction*. ANOVA allows for the testing of significance of individual parameters permitting the effect of crossover and mutation to

be statistically demonstrated. For issues which have been raised in the literature such as naive evolution [8], [9], ANOVA provides evidence which may or may not support the inclusion of the crossover parameter.

In addition, ANOVA allows for the testing of interaction between parameters. Interaction is simply the failure of one parameter to produce the same effect on the response variable at different levels of another parameter [19]. Examining interaction is important because a significant interaction means the effect of each parameter cannot be considered independently of the others. The interaction parameter is created by multiplying the crossover parameter by the mutation parameter and adding this parameter to the ANOVA model.

4) *Response curve analysis*. In ANOVA, once a parameter is demonstrated to be statistically significant the effect of the parameter may be modeled through an appropriate polynomial. Statistical testing can be carried out to assess if the shape of the response curve is predominantly linear or is comprised of higher order polynomials by partitioning the total variation of each parameter into its orthogonal polynomial contrast terms.

   Once the shape of the response curve is established, polynomial regression can be carried out to obtain estimates of the coefficients of the various parameters in the response curve equation. Importantly, if the interaction parameter is significant in the ANOVA model, then the overall equation must be found. If not, then the equations for crossover and mutation can be obtained separately.

   For fitted response curves which are comprised of quadratic or higher components, we can obtain the derivatives and find the values where the derivatives equal zero which yield estimates of the best value for each parameter. Additionally, confidence intervals can be calculated.

   However, if the fitted response curve is linear then a negative coefficient will correspond solely to a best rate of 100%, while a positive coefficient will correspond solely to a best rate of 0% since the minimum of a straight line can only occur at either end.

5) *Power*. The calculation of power for ANOVA can be made by using the effect size index $\boldsymbol{f}$ as described by Cohen [20].

6) *Availability*. ANOVA and regression are standard statistical models available in virtually all statistical software packages which are used on desktop computers.

7) *Applicability*. Randomized complete block design can be applied to other adaptive algorithms with little difficulty. It simply requires that the seeds, or any other sources of noise, are kept identical within each replicate so that the source can be blocked.

The GA was implemented in Java [21]. Statistical analysis was carried out using S-PLUS [22]. Power calculations were carried out using GPOWER [23].

A number of aspects of the analysis are discussed in more detail next.

D. Choice of Level of Significance

There are two types of errors associated with statistical testing. A type I error is the rejection of the null hypothesis when it is true. A type II error is the nonrejection of the null hypothesis when the alternative hypothesis is true. The probability of making a type I error is denoted by α and the probability of a type II error is denoted by β. Since the null hypothesis represents the most conservative proposal it is considered that a type I error is more serious than a type II error [18]. Thus, α is generally and arbitrarily set at a low level. This *level of significance* is traditionally set at values such as 10%, 5%, or 1%.

For published research a level of significance of 1% is often used [24]. P-values less than 1% suggest that the null hypothesis is *strongly rejected* or that the result is *highly statistically significant* [18]. In the present paper, we have employed 1% as our level of significance and correspondingly calculated 99% confidence intervals.

E. Level of Significance for Orthogonal Simultaneous Multiple Comparisons

In a situation of orthogonal simultaneous multiple comparisons within a parameter, it is necessary to modify the level of significance. This is because the probability of achieving one or more statistically significant results in n simultaneous multiple comparisons will exceed the level of significance chosen (1% in the present study). This is illustrated in (5)

$$P(\text{at least one significant result in } n) = 1 - (1 - \alpha)^n. \quad (5)$$

This occurs in ANOVA when the sum of squares for each parameter is partitioned into orthogonal contrast terms. In order to ensure that the probability of achieving one or more statistically significant results in n simultaneous multiple comparisons is *exactly* 1%, we use a modified level of significance for testing each of n orthogonal polynomial contrast terms calculated in accordance with (6)

$$\text{Modified level of significance} = 1 - (1 - \alpha)^{1/n}. \quad (6)$$

Our approach is different from the Bonferroni method [22] which, for the present work, would simply divide the overall level of significance by the number of simultaneous multiple comparisons. The Bonferroni method will ensure that the probability of achieving one or more statistically significant results in n simultaneous multiple comparisons is *no greater than* 1%. Thus, it yields an upper bound such that the actual probability of achieving one or more statistically significant results in n simultaneous multiple comparisons may be much smaller.

F. Power

As $1 - \beta$ is the probability of rejecting the null hypothesis when it is false, this is known as the *power* of the test. A power of 80% ($\beta = 0.2$) when there is moderate departure from the null hypothesis is considered desirable by convention [20]. The value of β is related to sample size. A sample size that is too small will generally fail to produce a significant result, while a sample size that is too large may be difficult to analyze and

wastes resources. It is, therefore, necessary to have some means of calculating whether the size of the sample chosen has sufficient power.

In order to calculate power, it is necessary to specify the degree to which the null hypothesis is false. This is quantifiable as a specific nonzero value using the unit-less effect size indices d and $\boldsymbol{f}$ as described by Cohen [20]. For ANOVA, by convention, a *small* effect size is an $\boldsymbol{f}$ value of 0.10, a *medium* effect size is an $\boldsymbol{f}$ value of 0.25, and a *large* effect size is an $\boldsymbol{f}$ value of 0.40.

In the present paper, differences in a specified number of epochs were first converted to the effect size index d, where

$$d = \frac{\mu_{\max} - \mu_{\min}}{\sigma} \tag{7}$$

where $\mu_{\max}$ is the largest population mean of a parameter level, $\mu_{\min}$ is the smallest population mean of a parameter level, and σ is the population standard deviation.

This results in a unit-less number to index the degree of departure from the null hypothesis of the alternative hypothesis, or more simply, the effect size we wish to detect [20].

Next, the conversion from d to $\boldsymbol{f}$ for ANOVA requires a knowledge of the pattern of separation for all means for all k levels of the parameter. Patterns identified by Cohen [20] are the following.

1) Minimum variability: One mean at each end of d, the remaining $k-2$ means all at the midpoint.
2) Intermediate variability: The k means equally spaced over d.
3) Maximum variability: The means are all at the end points of d.

Tables are available for the conversion from d to $\boldsymbol{f}$ for each scenario. If the pattern of separation is unknown an inspection of these tables illustrates that the most conservative approach is to assume the minimum variability pattern which results in $\boldsymbol{f}$ being at its smallest. In this case, $\boldsymbol{f}$ is calculated as

$$\boldsymbol{f} = d\sqrt{\frac{1}{2k}}. \tag{8}$$

It should be noted that power may be calculated *a priori* or *post hoc*. If the population standard deviation is known from prior research one can calculate *a priori* the sample size required to confer a specified power. On the other hand, if the population standard deviation is unknown but can be estimated once the study is concluded then *post hoc* power calculations indicate the ability of the present sample size to detect specified effect sizes.

As the present study was exploratory in nature and *a priori* assumptions about the population standard deviation could not be made, we strictly adhered to *post hoc* calculations. Thus, *unless statistical significance had been already demonstrated in the ANOVA analysis for the interaction parameter*, we continued to increase sample size by a factor of 5. This was enacted until at least 80% power was achieved for detecting a difference of five epochs for the interaction between crossover and mutation. This is because $\boldsymbol{f}$ is smallest for the interaction parameter since k is greatest for this parameter.

As a final remark, in the present research, we choose to calculate power based upon the ability to detect a difference of at least five epochs as noted above. This number was chosen as it most closely approximated the difference in the number of epochs detectable for the simplest problem *F1*, if we had calculated power using an $\boldsymbol{f}$ of 0.4 (*large* effect).

G. Simultaneous Confidence Intervals for the Plotted Response Curve

Plotting mean performance against parameter levels provides an initial estimate of the shape of the response curve. However, the shape of the curve may be compromised if the sample size is insufficient. To gauge the reliability of the trend, 99% simultaneous confidence intervals about each mean can be calculated. The z value for calculating simultaneous confidence intervals for n levels of an individual parameter corresponds to the probability given by (9)

$$P_{z\ \text{value}} = 1 - \left(\frac{1 - 0.99^{1/n}}{2}\right). \tag{9}$$

Note that while confidence intervals tighten as sample size increases, showing increased confidence about the location of the *population mean*, there is still a great deal of randomness in each *individual* run.

H. Pooled Analysis Design

If large data-sets are required these may not be able to be analyzed when a parameter has too many levels resulting in the statistical software having to deal with too many and too large matrices. In order to address this issue, we devised a pooled analysis design for the present study as follows.

1) For each individual experiment, we calculated the mean of the performance measure for each combination of crossover and mutation.
2) These data from individual experiments were concatenated into a new *pooled* data file. The response variable was now the mean of the performance measure averaged over the number of replicates in the individual experiment. This results in a smaller error variance as the average of a number of observations is expected to be closer than a single observation to the population mean. Each individual experiment denoted one level of the block parameter.
3) Analysis was carried out in the same manner as for individual experiments.

I. Estimates of Best Values for Parameters

Once the coefficients are obtained from the polynomial regression model it is straightforward to obtain an estimate of the best value for the specified parameter by differentiating and solving the response curve equation. 99% confidence intervals are then calculated using Taylor's Expansion (δ method) [25].

J. Workup Procedures to Ensure a Balanced ANOVA Design

A balanced design for ANOVA occurs if no data are missing or censored (threshold is not reached during the run of the GA). This is desirable since it results in the test statistic being more robust to small departures from the assumption of equal variances for the number of treatments. In addition, the power of

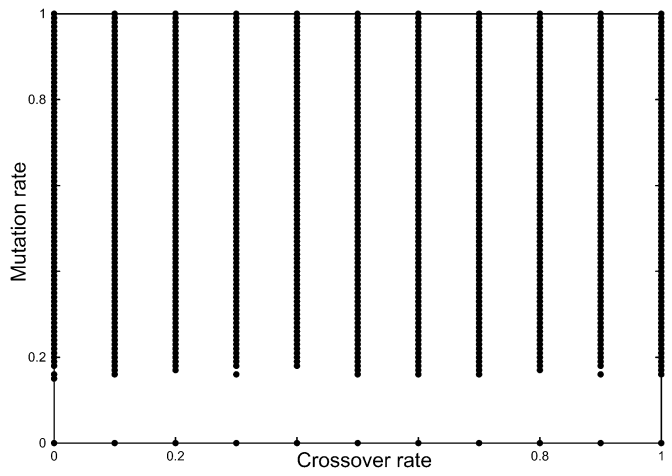


Fig. 1. Dot diagram for F1. Each dot represents an instance of censoring.

the ANOVA test is maximized. This was achieved by two consecutive workup procedures which were carried out for all four test functions.

1) Dot Diagrams: First, to minimize the occurrence of censoring in the present study a data-set of an arbitrary ten replicates was generated for all functions using crossover with values of zero to 1 with an interval of 0.1, and mutation with values of zero to 1 with an interval of 0.01. If on at least one occasion the threshold was not reached for a particular crossover rate and mutation rate combination, this was shown as a dot on the resultant dot diagram.

As illustrated in Fig. 1, for *F1* mutation rates of less than 0.15 and greater than zero were not associated with censoring. In contrast, all crossover rates from 0 to 1 were valid. Thus, at this point for *F1* the rates which could be considered to be reasonably free from censoring, so that the threshold value would be reached or exceeded on every run of the GA, were crossover rates of 0 to 1 with an interval of 0.1, and mutation rates of 0.01 to 0.14 with an interval of 0.01.

2) Finalizing Ranges for Exploratory Statistical Analysis: Second, to further ensure that no censored data would appear in the data-sets for analysis, and so finalize the ranges for exploratory statistical analysis to begin, we conducted the following exercise.

Using crossover and mutation rates not associated with censoring from the dot diagrams, an arbitrary ten data-sets of 100 replicates each were generated. Using S-PLUS, the combination of crossover rate and mutation rate resulting in the best performance was found in each data-set. When these ten combinations were collated, they demonstrated the lowest and highest rates of crossover and mutation associated with best performance. For *F1* crossover ranged from 0.8 to 1 and mutation ranged from 0.05 to 0.08.

However, to ensure that the ranges we would study could be considered robust, we allowed the ranges to widen one interval step on either side. Thus, as displayed in Table V, this made the finalized range for *F1* for crossover 0.7 to 1 with an interval of 0.1, and for mutation 0.04 to 0.09 with an interval of 0.01.

As a result of these two consecutive workup procedures, a balanced ANOVA design was achieved.

TABLE V
FINAL RANGES FOR CROSSOVER AND MUTATION

Test function	Crossover final range	Mutation final range
F1	0.7-1	0.04-0.09
F3	0.8-1	0.03-0.07
F2	0-0.7	0.18-0.24
F6	0-0.7	0.11-0.18

TABLE VI
F1-ANOVA OF 100 REPLICATES

Parameter	Df	Sum of Sq	Mean Sq	F Value	Pr(F)
Crossover	6	12347	2057.826	8.47756	0.0000000
Mutation	10	58701	5870.091	24.18282	0.0000000
Interaction	60	13664	227.733	0.93818	0.6117951
Block	99	51956	524.813	2.16205	0.0000000
Residuals	7524	1826361	242.738	-	-

Residual standard error: 15.58005, Estimated effects are balanced.

VI. RESULTS

A. Exploratory Analysis of Test Function F1

The results of analysis of data-sets containing 100 replicates, 500 replicates, and pooled results from five data-sets of 500 replicates are described consecutively to illustrate how statistics can be used to assist in exploratory analysis.

1) Results With 100 Replicates: Table VI displays ANOVA of 100 replicates.

Crossover and mutation were both highly statistically significant, while the interaction between crossover and mutation was not. *Post hoc* power calculations as shown in Table XVI show that while the power for detecting a difference of five epochs was greater than 97% for both crossover and mutation, the power for the interaction parameter was only 3.38%. Thus, the use of 100 replicates was too small to demonstrate statistical significance for interaction.

The response curve plots for crossover and mutation are displayed in Fig. 2(a) and (b).

While the response curve plot for mutation suggested a quadratic trend, the response curve plot for crossover was less obvious. Since only 100 replicates were used, the width of the simultaneous confidence intervals was very wide so that for crossover either a linear curve or a higher order polynomial such as a cubic curve could conceivably have fitted between the simultaneous confidence intervals. This is illustrated in Fig. 3(a) and (b).

As it is preferable to formally test for the shape of the response curve rather than relying on visual inspection, better information was obtained from the sum of squares partitioned into terms corresponding to orthogonal contrasts which represent polynomials. These data are shown in Table XXIII and suggested a linear trend for crossover and a quadratic trend for mutation.

However, given the lack of power associated with interaction it was necessary to repeat the analysis using an increased sample size. Adhering to our protocol of carrying out power calcula-

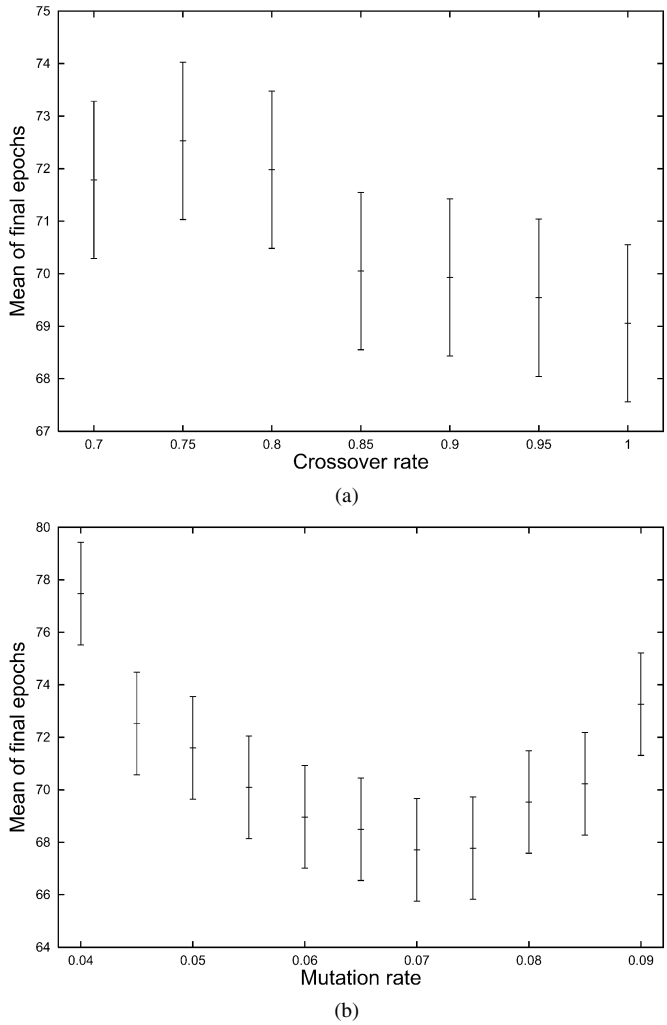


Fig. 2. (a) F1-Crossover response curve plot with 100 replicates. (b) F1-Mutation response curve plot with 100 replicates.

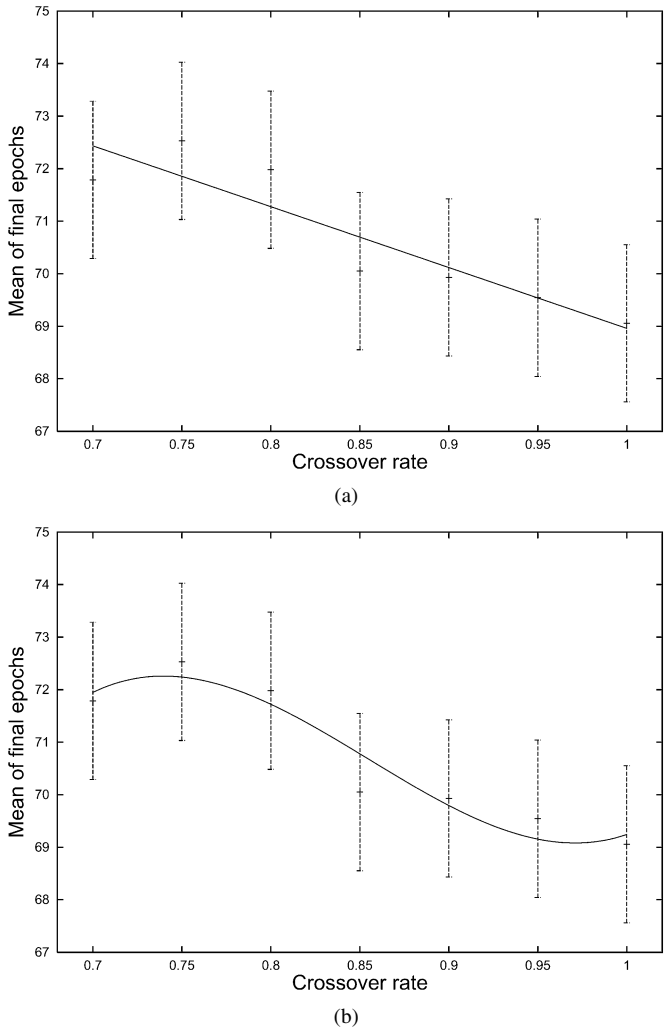


Fig. 3. (a) F1-Linear curve fitted through simultaneous confidence intervals. (b) F1-Cubic curve fitted through simultaneous confidence intervals.

tions on a strictly *post hoc* basis, we enacted a fivefold increase in the number of replicates.

2) Results With 500 Replicates: ANOVA of 500 replicates is shown in Table VII.

A similar pattern for the overall results was evident. That is, a highly significant result for crossover and mutation while a nonsignificant result for the interaction parameter.

Table XVII illustrates the improvement in power obtained by increasing the sample size though the power associated with the interaction parameter remained below the study threshold. The effect of increasing the number of replicates upon the width of the simultaneous confidence intervals for the response curves is shown in Fig. 4(a) and (b). The increase in the number of replicates reduced the width of the simultaneous confidence intervals producing clearer linear behavior for crossover and quadratic behavior for mutation. Both trends were affirmed in the partitioned sum of squares displayed in Table XXIV.

However, the continued lack of power associated with the interaction parameter meant that a further increase in the sample size was again required. We opted again for a fivefold increase in the number of replicates to 2500. However, this data-set could not be analyzed by S-PLUS due to the fact that the large number

TABLE VII
F1-ANOVA OF 500 REPLICATES

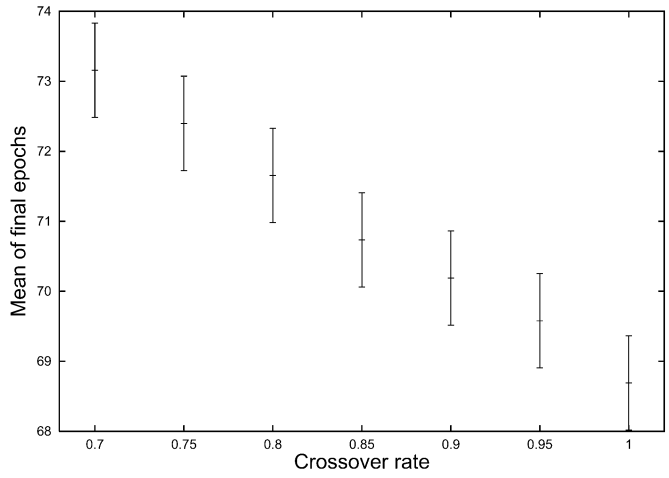
Parameter	Df	Sum of Sq	Mean Sq	F Value	Pr(F)
Crossover	6	82952	13825.38	56.20533	0.0000000
Mutation	10	208227	20822.75	84.65223	0.0000000
Interaction	60	12386	206.44	0.83925	0.8079445
Block	499	237465	475.88	1.93464	0.0000000
Residuals	37924	9328542	245.98	-	-

Residual standard error: 15.68375, Estimated effects are balanced.

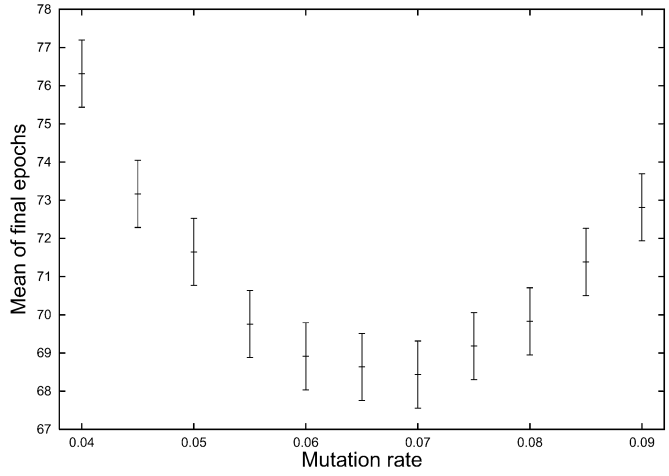
of levels for the block variable meant that the calculations involved too many and too large matrices. As such, the pooled analysis design was implemented.

3) Results of the Pooled Analysis: Table VIII shows ANOVA of the pooled data-set from five data-sets of 500 replicates. Both crossover and mutation were again highly statistically significant. However, the interaction between crossover and mutation was not with a p-value of 0.0377.

Post hoc power calculations are displayed in Table XVIII. The increase in replicates now resulted in 100% power to detect a difference of five epochs for the interaction parameter. As the

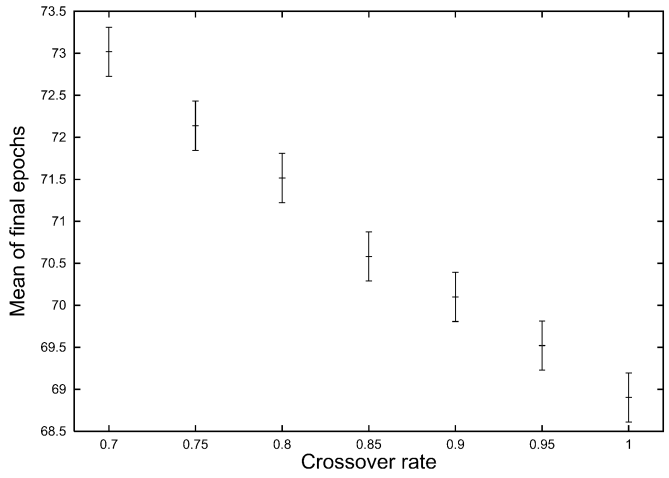


(a)

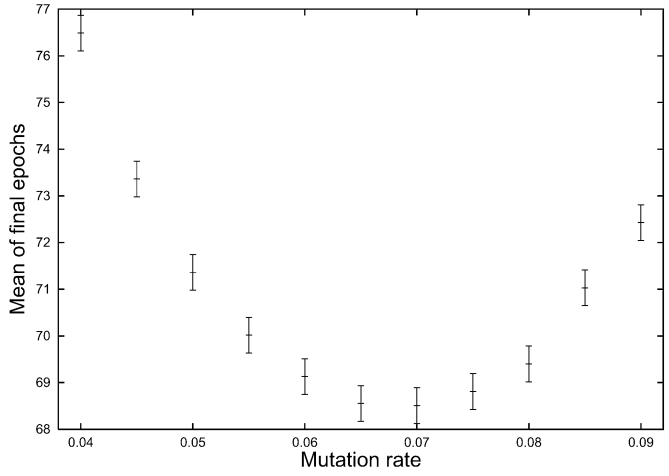


(b)

Fig. 4. (a) F1-Crossover response curve plot with 500 replicates. (b) F1-Mutation response curve plot with 500 replicates.



(a)



(b)

Fig. 5. (a) F1-Crossover response curve plot from pooled analysis. (b) F1-Mutation response curve plot from pooled analysis.

TABLE VIII
F1-POOLED ANOVA ANALYSIS

Parameter	Df	Sum of Sq	Mean Sq	F Value	Pr(F)
Crossover	6	714.601	119.1002	256.1305	0.0000000
Mutation	10	2153.876	215.3876	463.2010	0.0000000
Interaction	60	38.977	0.6496	1.3970	0.0377493
Block	4	1.381	0.3453	0.7426	0.5635587
Residuals	304	141.359	0.4650	-	-

Residual standard error: 0.6819076, Estimated effects are balanced.

power threshold of the study had been exceeded, it was not necessary to increase the sample size any further.

The response curve plots for crossover and mutation from the pooled analysis are displayed in Fig. 5(a) and (b). As can be seen the width of the simultaneous confidence intervals has been further tightened. The partitioned sum of squares shown in Table XXV illustrated strong agreement with the plots. However, for mutation a cubic effect was now significant though the quadratic effect remained predominant as evidenced when comparing the magnitude of the respective sum of squares.

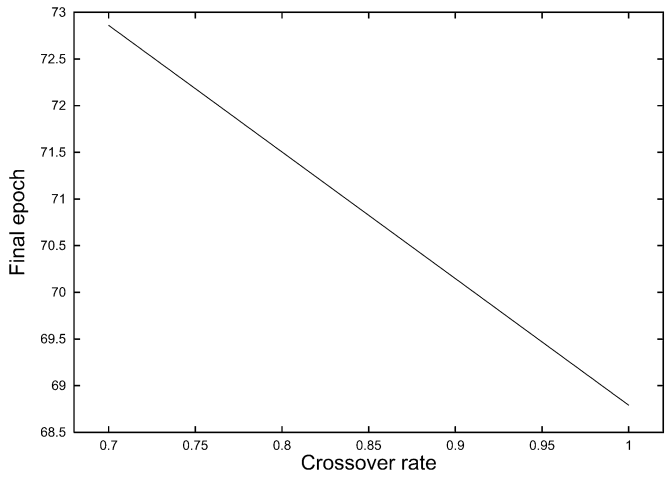
In conclusion, these data suggested that both crossover and mutation are highly important parameters in the GA for the *F1* problem domain. The behavior of crossover is linear, while the behavior of mutation is predominantly quadratic with some cubic component. The interaction observed between crossover and mutation is not significant and, therefore, is of little practical importance.

Using polynomial regression separate fitted response curves for crossover and mutation were obtained. These are illustrated in Fig. 6(a) and (b) and the equations are given in Table XXX. Using these equations the best values for crossover and mutation were calculated and the overall results are displayed in Table IX.

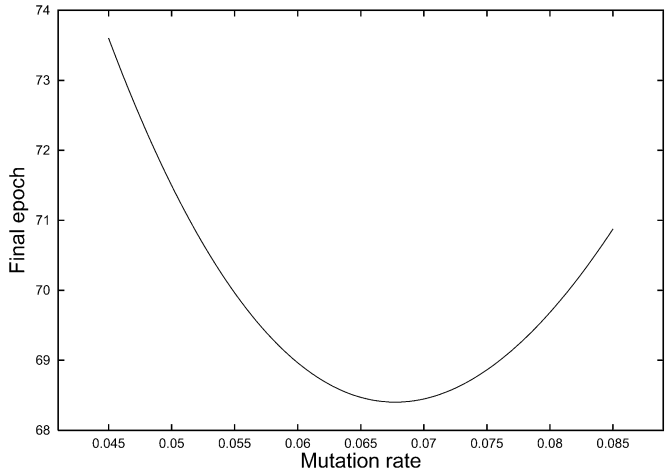
B. Exploratory Analysis of Test Function F3

ANOVA of the pooled data-set for *F3* is shown in Table X. Crossover and mutation were highly statistically significant, while the interaction between crossover and mutation was not. *Post hoc* power calculations displayed in Table XIX show that the power for detecting a difference of five epochs for the interaction parameter was 88.27%, exceeding the threshold for the present study. As such, there was no further need to increase the sample size.

An examination of the partitioned sum of squares shown in Table XXVI confirmed a linear trend for crossover and a quadratic trend for mutation. Using polynomial regression the fitted response curves for crossover and mutation were

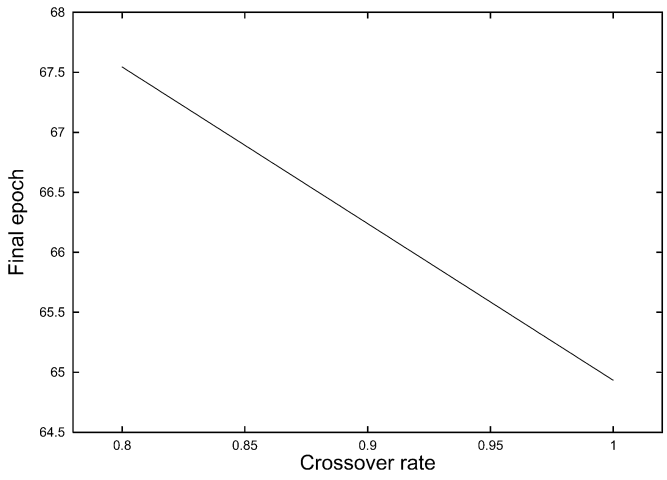


(a)

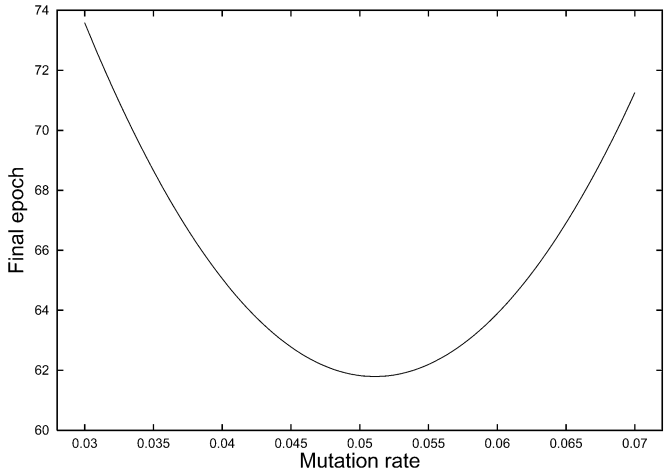


(b)

Fig. 6. (a) Fitted response curve: F1-crossover. (b) Fitted response curve: F1-mutation.



(a)



(b)

Fig. 7. (a) Fitted response curve: F3-crossover. (b) Fitted response curve: F3-mutation.

TABLE IX
F1-OVERALL RESULTS FOR CROSSOVER AND MUTATION

Parameter	Response curve shape	Estimated best value	99% CI
Crossover	Linear	100%	-
Mutation	Cubic	6.77%	6.60%-6.95%

TABLE X
F3-POOLED ANOVA ANALYSIS

Parameter	Df	Sum of Sq	Mean Sq	F Value	Pr(F)
Crossover	4	251.835	62.9588	51.8074	0.0000000
Mutation	8	3460.606	432.5757	355.9567	0.0000000
Interaction	32	50.045	1.5639	1.2869	0.1550913
Block	4	12.390	3.0974	2.5488	0.0409906
Residuals	176	213.884	1.2152	-	-

Residual standard error: 1.102383, Estimated effects are balanced.

obtained. These are illustrated in Fig. 7(a) and (b) and the equations given in Table XXX. Using these equations the best values for crossover and mutation were calculated and the overall results are displayed in Table XI.

TABLE XI
F3-OVERALL RESULTS FOR CROSSOVER AND MUTATION

Parameter	Response curve shape	Estimated best value	99% CI
Crossover	Linear	100%	-
Mutation	Quadratic	5.11%	5.07%-5.15%

TABLE XII
F2-POOLED ANOVA ANALYSIS

Parameter	Df	Sum of Sq	Mean Sq	F Value	Pr(F)
Crossover	14	29291.3	2092.235	46.1088	0.000000000
Mutation	12	103575.8	8631.317	190.2173	0.000000000
Interaction	168	10717.5	63.795	1.4059	0.001550061
Block	4	820.0	205.006	4.5179	0.001298162
Residuals	776	35211.8	45.376	-	-

Residual standard error: 6.736177, Estimated effects are balanced.

C. Exploratory Analysis of Test Function F2

1) Results of the Pooled Analysis: Table XII shows ANOVA analysis of the pooled data-set for *F2*.

Crossover and mutation were highly statistically significant as was the interaction between crossover and mutation with

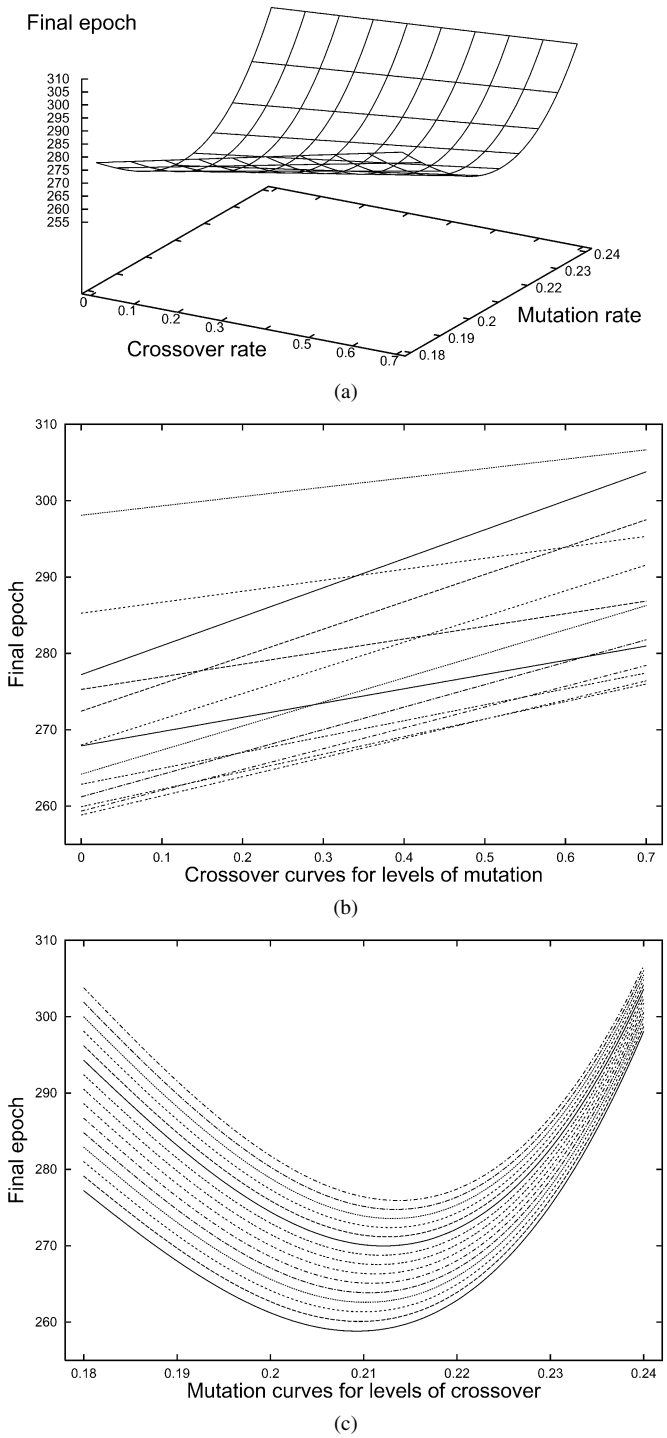


Fig. 8. (a) Fitted response curve: F2. (b) Fitted response curve: F2-crossover. (c) Fitted response curve: F2-mutation.

a p-value of 0.00155. Since the interaction parameter demonstrated strong statistical significance, no further increments in sample size were necessary.

Examination of the sum of squares partitioned into orthogonal polynomial contrast terms as shown in Table XXVII suggested a linear trend for crossover and a cubic trend for mutation with the predominant effect for the latter arising from the quadratic term. Partitioning of the sum of squares of the interaction parameter showed only a statistically significant effect (p-value less than 0.01) for the linear:linear term (that is, the linear component of crossover multiplied by the linear component of mutation).

TABLE XIII
F2-OVERALL RESULTS FOR CROSSOVER AND MUTATION

Parameter	Response curve shape	Estimated best value	99% CI
Crossover	Linear	0%	-
Mutation	Cubic	21.15%	21.01%-21.30%
Interaction	Linear:Linear	-	-

TABLE XIV
F6-POOLED ANOVA ANALYSIS

Parameter	Df	Sum of Sq	Mean Sq	F Value	Pr(F)
Crossover	14	54420.8	3887.20	93.4536	0.0000000
Mutation	14	162014.1	11572.44	278.2172	0.0000000
Interaction	196	50461.5	257.46	6.1896	0.0000000
Block	4	77.3	19.31	0.4643	0.7619715
Residuals	896	37269.1	41.59	-	-

Residual standard error: 6.449417, Estimated effects are balanced.

As the interaction parameter was found to be significant, in contrast to the results for *F1* and *F3*, polynomial regression was used to obtain the overall three-dimensional (3-D) equation for the response curve and this is given in Table XXX. Fig. 8(a) illustrates this overall 3-D response curve and Fig. 8(b) and (c) illustrates two-dimensional (2-D) slices corresponding to crossover and mutation, respectively.

Fig. 8(b) illustrates consistent positive slopes for the crossover curves indicating a worsening of performance as the crossover rate increased. Additionally, it should be noted that the top curve and the second curve from the top correspond to mutation values of 24% and 18%, respectively. As the other curves fall inside these extremes, this illustrates how this cross section actually curves into the page. In Fig. 8(c), we see the curved trend of each mutation curve. In this graph, the top curve corresponds to a crossover rate of 70% and the bottom curve corresponds to a crossover rate of 0%. This suggests that mutation performs best when the crossover rate is 0%.

Using the equation where the rate of crossover was 0% the best value for mutation was calculated. The overall results of the analysis are shown in Table XIII.

D. Exploratory Analysis of Test Function F6

1) Results of the Pooled Analysis: Table XIV shows ANOVA analysis of the pooled data-set for *F6*.

Paralleling the results for *F2*, both crossover and mutation were highly statistically significant together with the interaction. As before, strong statistical significance for the interaction parameter meant that no further increments in sample size were necessary.

Inspection of the sum of squares partitioned into orthogonal polynomial contrast terms as shown in Table XXVIII demonstrated up to quadratic behavior for crossover with the linear component being predominant, while for mutation up to cubic behavior with the quadratic effect being predominant. Interaction was more complex than for *F2* with significant interaction terms: linear:linear, quadratic:linear, linear:quadratic, and linear:cubic.

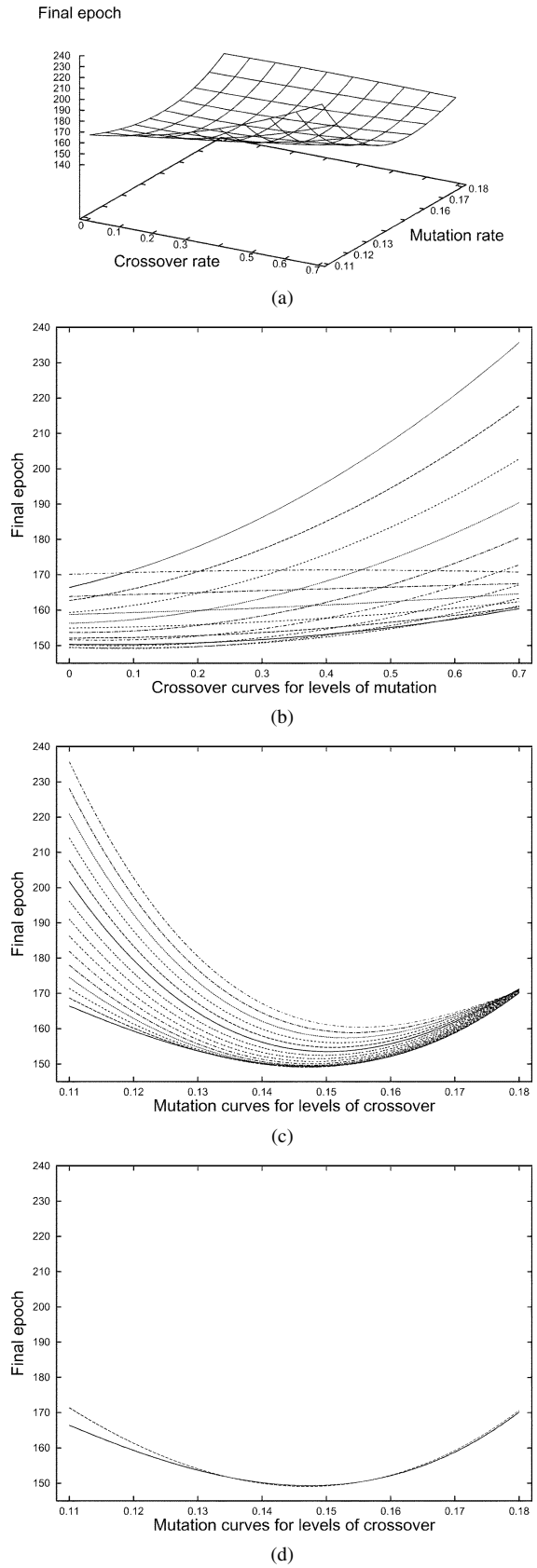


Fig. 9. (a) Fitted response curve: F6. (b) Fitted response curve: F6-crossover. (c) Fitted response curve: F6-mutation. (c) Fitted response curves for crossover 0% and 10%: F6-mutation.

Again, using polynomial regression the overall 3-D equation for the response curve was obtained and is given in Table XXX.

TABLE XV
F6-OVERALL RESULTS FOR CROSSOVER AND MUTATION

Parameter	Response curve shape	Estimated best value	99% CI
Crossover	Quadratic	0%	-
Mutation	Cubic	15.01%	14.80%-15.22%
Interaction	Linear:Linear	-	-
	Quadratic:Linear	-	-
	Linear:Quadratic	-	-
	Linear:Cubic	-	-

Fig. 9(a) illustrates the overall 3-D response curve and Fig. 9(b) and (c) illustrates 2-D slices corresponding to crossover and mutation, respectively.

In Fig. 9(c), we see the curved trend of each mutation curve. However, Fig. 9(d), which displays mutation curves for crossover rates of 0% and 10%, respectively, illustrates that performance was predicted to improve very slightly with the latter crossover rate. This was also seen when examining mutation rates for crossover rates of 5% and 15%. However, to assess in a practical fashion if these differences would be apparent in a data-set focusing upon this range, we generated five 500 replicate data-sets keeping the mutation range the same but narrowing the range of crossover from 0% to 15% inclusive with an interval of 1%.

As shown in Table XXIX ANOVA analysis illustrated that the differences in performance due to crossover over this range were marginal with a p-value of 0.0208 despite the power being high at 91.63%. Moreover, the partitioned sum of squares illustrated that the effect of crossover was solely linear with a p-value of 0.0003. Regression analysis confirmed that the coefficient for the linear term was positive indicating a worsening of performance as the crossover rate increased.

Thus, using the equation where the rate of crossover was 0% the best value for mutation was calculated. The overall results of the analysis are shown in Table XV.

VII. DISCUSSION

GAs have been studied in computer science and used in real-world applications to find solutions to difficult problems. However, there is no generally accepted methodology to assess which parameters significantly affect performance, whether these parameters interact and how performance varies with respect to changes in parameters. This study describes a statistical methodology for the exploratory study of genetic and other adaptive algorithms addressing these issues.

Generically, once the algorithm and the problem domain have been specified, the steps in the analysis are the following.

1) Identify sources of variation and modify the algorithm to generate blocked runs.
2) Use a workup procedure to minimize the appearance of censored observations and to finalize starting ranges for parameters.
3) Generate an initial data-set consisting of an arbitrary number of replicates. Typically, we have found 100 replicates to be a useful starting point.

TABLE XVI
F1-POWER WITH 100 REPLICATES

Parameter	Difference (epochs)	Effect size index f	Power
Crossover	10	0.17154	100%
Crossover	5	0.08578	99.99%
Crossover	3	0.05146	84.11%
Crossover	2	0.03431	35.36%
Crossover	1	0.01715	5.19%
Crossover	Large	0.4	100%
Crossover	Medium	0.25	100%
Crossover	Small	0.1	100%
Mutation	10	0.13684	100%
Mutation	5	0.06842	97.84%
Mutation	3	0.04105	44.53%
Mutation	2	0.02737	13.03%
Mutation	1	0.01368	2.57%
Mutation	Large	0.4	100%
Mutation	Medium	0.25	100%
Mutation	Small	0.1	100%
Interaction	10	0.05172	27.58%
Interaction	5	0.02586	**3.38%**
Interaction	3	0.01552	1.62%
Interaction	2	0.01034	1.25%
Interaction	1	0.00517	1.06%
Interaction	Large	0.4	100%
Interaction	Medium	0.25	100%
Interaction	Small	0.1	99.52%

Mean square error = 15.58005 epochs.

TABLE XVII
F1-POWER WITH 500 REPLICATES

Parameter	Difference (epochs)	Effect size index f	Power
Crossover	10	0.17041	100%
Crossover	5	0.08520	100%
Crossover	3	0.05112	100%
Crossover	2	0.03408	>99.37%
Crossover	1	0.01704	>36.65%
Crossover	Large	0.4	100%
Crossover	Medium	0.25	100%
Crossover	Small	0.1	100%
Mutation	10	0.13594	100%
Mutation	5	0.06797	100%
Mutation	3	0.04078	>99.94%
Mutation	2	0.02719	>83.66%
Mutation	1	0.01359	>13.55%
Mutation	Large	0.4	100%
Mutation	Medium	0.25	100%
Mutation	Small	0.1	100%
Interaction	10	0.05138	>99.84%
Interaction	5	0.02569	**>29.06%**
Interaction	3	0.01541	>5.40%
Interaction	2	0.01028	>2.33%
Interaction	1	0.00514	>1.26%
Interaction	Large	0.4	100%
Interaction	Medium	0.25	100%
Interaction	Small	0.1	100%

Mean square error = 15.68375 epochs.
Note: GPOWER can only accept sample sizes of up to 32000.
The sample size for 500 replicates was 38500.
Thus, where a > symbol is used power was calculated
using a sample size of 32000 while the actual power
would be marginally greater.

4) Calculate power *post hoc* based upon a chosen effect size. If at least 80% power is not achieved increase the sample size.
5) Conduct (pooled) ANOVA analysis and determine which parameters are statistically significant.
6) For parameters which are statistically significant, partition the sum of squares into polynomial contrast terms. Determine which polynomial terms are statistically significant.
7) Use polynomial regression to obtain the coefficients for the overall response curve (if the interaction parameter is statistically significant) or to obtain the coefficients for the response curve for each parameter separately (if the interaction parameter is not statistically significant).
8) Differentiate and solve the response curve for each parameter to obtain best values and calculate confidence intervals.

Before discussing the specific results of our study it should be prefaced that the present research aimed to provide a statistical methodology by demonstrating its practical use in well known test functions. In this regard, the number of parameters and the suite of problems is restricted. Further research using a statistical approach with an expanded set of parameters, in both continuous and discrete problem domains, will be necessary to expand upon these initial findings.

The analysis of *F1* illustrates the way in which our methodology was used to make informed decisions when exploring the relationship between crossover and mutation on a specified problem. Initially, workup procedures yielded starting ranges for crossover and mutation. ANOVA analysis of an initial data-set of 100 replicates demonstrated a statistically significant effect upon performance of both crossover and mutation with nonsignificance for the interaction parameter.

TABLE XVIII
F1-POWER OF THE POOLED ANALYSIS

Parameter	Difference (epochs)	Effect size index f	Power
Crossover	10	3.9193	100%
Crossover	5	1.9597	100%
Crossover	3	1.1758	100%
Crossover	2	0.78386	100%
Crossover	1	0.39193	100%
Crossover	Large	0.4	100%
Crossover	Medium	0.25	90.39%
Crossover	Small	0.1	9.83%
Mutation	10	3.1265	100%
Mutation	5	1.5633	100%
Mutation	3	0.93796	100%
Mutation	2	0.62531	100%
Mutation	1	0.31265	97.94%
Mutation	Large	0.4	99.99%
Mutation	Medium	0.25	82.55%
Mutation	Small	0.1	6.96%
Interaction	10	1.1817	100%
Interaction	5	0.59086	**100%**
Interaction	3	0.35452	79.01%
Interaction	2	0.23634	23.79%
Interaction	1	0.11817	3.11%
Interaction	Large	0.4	92.65%
Interaction	Medium	0.25	29.05%
Interaction	Small	0.1	2.33%

Mean square error = 0.6819076 epochs.

Attempting to gauge the shape of the response curve plots was compromised by the small sample size. As seen, the width of the simultaneous 99% confidence intervals made it unclear as to whether the trend for crossover was linear or included higher order components.

TABLE XIX
F3-POWER OF THE POOLED ANALYSIS

Parameter	Difference (epochs)	Effect size index f	Power
Crossover	10	2.6652	100%
Crossover	5	1.3326	100%
Crossover	3	0.79956	100%
Crossover	2	0.53304	100%
Crossover	1	0.26652	75.25%
Crossover	Large	0.4	99.49%
Crossover	Medium	0.25	67.45%
Crossover	Small	0.1	6.26%
Mutation	10	1.9865	100%
Mutation	5	0.99327	100%
Mutation	3	0.59596	100%
Mutation	2	0.39731	97.74%
Mutation	1	0.19865	26.92%
Mutation	Large	0.40	97.93%
Mutation	Medium	0.25	51.41%
Mutation	Small	0.1	4.12%
Interaction	10	0.88840	100%
Interaction	5	0.44420	**88.27%**
Interaction	3	0.26652	23.21%
Interaction	2	0.17768	6.34%
Interaction	1	0.08884	1.76%
Interaction	Large	0.4	75.30%
Interaction	Medium	0.25	18.64%
Interaction	Small	0.1	2.02%

Mean square error = 1.1865 epochs.

TABLE XX
F2-POWER OF THE POOLED ANALYSIS

Parameter	Difference (epochs)	Effect size index f	Power
Crossover	10	0.27104	100%
Crossover	5	0.13552	56.28%
Crossover	3	0.08131	11.87%
Crossover	2	0.05421	4.05%
Crossover	1	0.02710	1.53%
Crossover	Large	0.4	100%
Crossover	Medium	0.25	99.96%
Crossover	Small	0.1	22.88%
Mutation	10	0.29113	100%
Mutation	5	0.14557	70.38%
Mutation	3	0.08734	16.61%
Mutation	2	0.05823	5.24%
Mutation	1	0.02911	1.69%
Mutation	Large	0.40	100%
Mutation	Medium	0.25	99.98%
Mutation	Small	0.1	25.48%
Interaction	10	0.07517	2.04%
Interaction	5	0.03759	**1.21%**
Interaction	3	0.02255	1.07%
Interaction	2	0.01503	1.03%
Interaction	1	0.00752	1.01%
Interaction	Large	0.4	99.97%
Interaction	Medium	0.25	62.57%
Interaction	Small	0.1	3.32%

Mean square error = 6.736177 epochs.

In contrast, the sum of squares partitioned into terms corresponding to orthogonal polynomial contrasts demonstrated predominantly linear and quadratic trends for crossover and mutation, respectively. Although this dispelled the ambiguity associated with the data obtained from visual inspection, the subsequent power calculations clearly showed a lack of power for the interaction parameter. Therefore, increases in sample size were required. This was carried out until the appropriate power for the interaction parameter was achieved. At this point, polynomial regression was used to obtain fitted response curves and best values with 99% confidence intervals were calculated.

TABLE XXI
F6-POWER OF THE POOLED ANALYSIS

Parameter	Difference (epochs)	Effect size index f	Power
Crossover	10	.28308	100%
Crossover	5	.14154	72.65%
Crossover	3	.08492	17.11%
Crossover	2	.05661	5.30%
Crossover	1	.02830	1.69%
Crossover	Large	.4	100%
Crossover	Medium	.25	99.99%
Crossover	Small	.1	28.86%
Mutation	10	.28308	100%
Mutation	5	.14154	72.65%
Mutation	3	.08492	17.11%
Mutation	2	.05661	5.30%
Mutation	1	.02830	1.69%
Mutation	Large	.4	100%
Mutation	Medium	.25	99.99%
Mutation	Small	.1	28.86%
Interaction	10	.07309	2.05%
Interaction	5	.03654	**1.21%**
Interaction	3	.02192	1.07%
Interaction	2	.01461	1.03%
Interaction	1	.00730	1.01%
Interaction	Large	0.4	99.99%
Interaction	Medium	0.25	69.01%
Interaction	Small	0.1	3.56%

Mean square error = 6.449417 epochs.

TABLE XXII
F6-POWER OF THE POOLED ANALYSIS FOR CROSSOVER 0% TO 15%

Parameter	Difference (epochs)	Effect size index f	Power
Crossover	10	.32905	100%
Crossover	5	.16452	**91.63%**
Crossover	3	.09871	29.32%
Crossover	2	.06581	8.24%
Crossover	1	.03290	2.02%
Crossover	Large	.4	100%
Crossover	Medium	.25	100%
Crossover	Small	.1	30.54%

Mean square error = 5.372283 epochs.

TABLE XXIII
F1-PARTITIONED SUM OF SQUARES WITH 100 REPLICATES

Parameter	Df	Sum of Sq	Mean Sq	F Value	Pr(F)
Crossover	6	12347	2057.83	8.4776	**0.0000000**
Crossover adjusted level of significance = 0.001673654					
Power of 1	1	10330	10329.82	42.5554	**0.0000000**
Power of 2	1	38	38.13	0.1571	0.6918712
Power of 3	1	976	975.98	4.0207	0.0449809
Power of 4	1	681	680.92	2.8052	0.0940032
Power of 5	1	14	13.70	0.0564	0.8122398
Power of 6	1	308	308.41	1.2705	0.2597008
Mutation	10	58701	5870.09	24.1828	**0.0000000**
Mutation adjusted level of significance = 0.001004529					
Power of 1	1	11389	11388.70	46.9176	**0.0000000**
Power of 2	1	44725	44724.56	184.2503	**0.0000000**
Power of 3	1	2	2.16	0.0089	0.9248439
Power of 4	1	1069	1068.68	4.4026	0.0359176
Power of 5	1	553	552.87	2.2776	0.1312950
Power of 6	1	452	451.55	1.8602	0.1726404
Power of 7	1	2	1.66	0.0068	0.9340925
Power of 8	1	487	486.78	2.0054	0.1567837
Power of 9	1	20	20.44	0.0842	0.7717104
Power of 10	1	4	3.52	0.0145	0.9041185

TABLE XXIV
F1-PARTITIONED SUM OF SQUARES WITH 500 REPLICATES

Parameter	Df	Sum of Sq	Mean Sq	F Value	Pr(F)
Crossover	6	82952	13825.4	56.2053	**0.0000000**
Crossover adjusted level of significance = 0.001673654					
Power of 1	1	82662	82661.9	336.0514	**0.0000000**
Power of 2	1	40	39.8	0.1619	0.6874415
Power of 3	1	31	31.2	0.1267	0.7219155
Power of 4	1	150	150.4	0.6116	0.4341996
Power of 5	1	17	16.5	0.0672	0.7954938
Power of 6	1	52	52.5	0.2132	0.6442386
Mutation	10	208227	20822.7	84.6522	**0.0000000**
Mutation adjusted level of significance = 0.001004529					
Power of 1	1	32019	32018.7	130.1681	**0.0000000**
Power of 2	1	174262	174261.6	708.4383	**0.0000000**
Power of 3	1	959	959.3	3.9000	0.0482925
Power of 4	1	10	10.1	0.0409	0.8398032
Power of 5	1	108	107.8	0.4381	0.5080262
Power of 6	1	29	28.6	0.1162	0.7331794
Power of 7	1	350	349.8	1.4219	0.2330996
Power of 8	1	90	90.1	0.3663	0.5450536
Power of 9	1	344	344.1	1.3989	0.2369111
Power of 10	1	57	57.4	0.2335	0.6289593

TABLE XXV
F1-PARTITIONED SUM OF SQUARES OF POOLED ANALYSIS

Parameter	Df	Sum of Sq	Mean Sq	F Value	Pr(F)
Crossover	6	714.601	119.100	256.130	**0.0000000**
Crossover adjusted level of significance = 0.001673654					
Power of 1	1	708.852	708.852	1524.420	**0.0000000**
Power of 2	1	3.884	3.884	8.352	0.0041303
Power of 3	1	0.065	0.065	0.140	0.7082399
Power of 4	1	0.199	0.199	0.429	0.5131917
Power of 5	1	0.344	0.344	0.740	0.3904751
Power of 6	1	1.257	1.257	2.703	0.1011870
Mutation	10	2153.876	215.388	463.201	**0.0000000**
Mutation adjusted level of significance = 0.001004529					
Power of 1	1	473.173	473.173	1017.581	**0.0000000**
Power of 2	1	1665.259	1665.259	3581.217	**0.0000000**
Power of 3	1	6.476	6.476	13.926	**0.0002269**
Power of 4	1	3.828	3.828	8.232	0.0044039
Power of 5	1	2.830	2.830	6.087	0.0141682
Power of 6	1	0.397	0.397	0.854	0.3560224
Power of 7	1	0.984	0.984	2.116	0.1467925
Power of 8	1	0.760	0.760	1.634	0.2021186
Power of 9	1	0.154	0.154	0.330	0.5658050
Power of 10	1	0.015	0.015	0.031	0.8595995

TABLE XXVI
F3-PARTITIONED SUM OF SQUARES OF POOLED ANALYSIS

Parameter	Df	Sum of Sq	Mean Sq	F Value	Pr(F)
Crossover	4	196.365	49.091	34.871	**0.0000000**
Crossover adjusted level of significance = 0.00250943					
Power of 1	1	191.806	191.806	136.247	**0.0000000**
Power of 2	1	0.773	0.773	0.549	0.4596335
Power of 3	1	1.118	1.118	0.794	0.3740606
Power of 4	1	2.668	2.668	1.895	0.1703326
Mutation	8	3520.036	440.004	312.551	**0.0000000**
Mutation adjusted level of significance = 0.001255503					
Power of 1	1	127.126	127.126	90.302	**0.0000000**
Power of 2	1	3377.901	3377.901	2399.447	**0.0000000**
Power of 3	1	6.795	6.795	4.827	0.0293291
Power of 4	1	4.257	4.257	3.024	0.0837819
Power of 5	1	2.047	2.047	1.454	0.2294650

TABLE XXVII
F2-PARTITIONED SUM OF SQUARES OF POOLED ANALYSIS

Parameter	Df	Sum of Sq	Mean Sq	F Value	Pr(F)
Crossover	14	29291.3	2092.24	46.109	**0.0000000**
Crossover adjusted level of significance = 0.0007176235					
Power of 1	1	28663.0	28662.98	631.676	**0.0000000**
Power of 2	1	149.4	149.43	3.293	0.0699523
Power of 3	1	60.2	60.24	1.328	0.2495765
Power of 4	1	62.7	62.66	1.381	0.2403146
Power of 5	1	0.1	0.07	0.002	0.9677584
Power of 6	1	96.2	96.19	2.120	0.1458023
Power of 7	1	5.3	5.33	0.118	0.7318478
Power of 8	1	64.0	64.01	1.411	0.2353115
Power of 9	1	30.2	30.15	0.665	0.4152246
Power of 10	1	73.4	73.37	1.617	0.2039037
Power of 11	1	27.2	27.20	0.599	0.4390594
Power of 12	1	12.3	12.28	0.271	0.6030844
Power of 13	1	43.8	43.83	0.966	0.3259990
Power of 14	1	3.5	3.54	0.078	0.7799435
Mutation	12	103575.8	8631.32	190.217	**0.0000000**
Mutation adjusted level of significance = 0.0008371774					
Power of 1	1	3878.8	3878.80	85.481	**0.0000000**
Power of 2	1	96213.2	96213.19	2120.350	**0.0000000**
Power of 3	1	2662.8	2662.77	58.682	**0.0000000**
Power of 4	1	20.8	20.84	0.459	0.4982083
Power of 5	1	13.5	13.46	0.297	0.5862050
Power of 6	1	172.7	172.68	3.805	0.0514453
Power of 7	1	5.3	5.31	0.117	0.7323648
Power of 8	1	72.0	72.03	1.587	0.2080834
Power of 9	1	116.6	116.57	2.569	0.1093895
Power of 10	1	57.4	57.37	1.264	0.2611975
Power of 11	1	343.5	343.54	7.571	0.0060701
Power of 12	1	19.3	19.26	0.424	0.5149314
Interaction	168	10717.5	63.79	1.406	**0.0015501**
Interaction adjusted level of significance = 0.00005982164. *Only significant results shown.*					
Power of 1: Power of 1	1	2924.0	2923.96	64.438	**0.0000000**

Looking at the results from the suite of test functions together, crossover appears to have a predominantly linear effect upon performance. For *F1* and *F3* the positive gradient suggests selecting a rate as high as possible, while for *F2* and *F6* the negative gradient suggests its possible exclusion. As noted earlier, Schaffer *et al.* [8] documented a relative insensitivity to crossover for these same functions and our research adds to evidence supporting the effectiveness of naive evolution for certain problems. Indeed, as suggested earlier, naive evolution may be a powerful search algorithm in its own right as subtly commented by Eshelman [9]. Given that our study has controlled for the effect of seed, we may be obtaining a clearer perspective of the actual behavior of crossover than has been seen previously. Whatever the case, the observation in our work that crossover appears predominantly linear and that the direction of its slope is problem specific is certainly of practical interest. It may be possible to correlate this behavior with particular classes of problems making it easier to decide how to make the best use of the crossover parameter. We are currently investigating this idea further.

In contrast, mutation appears to have a consistent and predominantly quadratic effect upon performance. Why the effect should be more complex than that of crossover is another question of interest as it may lead to further insights into GA dynamics. The best values of mutation range from 5.11% to 20.92% (corresponding to a *bit-flipping mutation rate* of up to approximately 10%). These mutation rates add to a growing

TABLE XXVIII
F6-PARTITIONED SUM OF SQUARES OF POOLED ANALYSIS

Parameter	Df	Sum of Sq	Mean Sq	F Value	Pr(F)
Crossover	14	54420.8	3887.2	93.454	**0.0000000**
Crossover adjusted level of significance = 0.0007176235					
Power of 1	1	51558.8	51558.8	1239.544	**0.0000000**
Power of 2	1	2723.0	2723.0	65.465	**0.0000000**
Power of 3	1	0.1	0.1	0.002	0.9672032
Power of 4	1	0.2	0.2	0.005	0.9438726
Power of 5	1	14.2	14.2	0.340	0.5597281
Power of 6	1	10.2	10.2	0.246	0.6203542
Power of 7	1	5.0	5.0	0.121	0.7282759
Power of 8	1	17.3	17.3	0.417	0.5187929
Power of 9	1	59.5	59.5	1.430	0.2321141
Power of 10	1	1.7	1.7	0.040	0.8419240
Power of 11	1	0.0	0.0	0.000	0.9855772
Power of 12	1	0.1	0.1	0.002	0.9613900
Power of 13	1	30.7	30.7	0.739	0.3901418
Power of 14	1	0.0	0.0	0.000	0.9893777
Mutation	14	162014.1	11572.4	278.217	**0.0000000**
Mutation adjusted level of significance = 0.0007176235					
Power of 1	1	49729.9	49729.9	1195.574	**0.0000000**
Power of 2	1	111146.3	111146.3	2672.109	**0.0000000**
Power of 3	1	485.9	485.9	11.681	**0.0006599**
Power of 4	1	209.9	209.9	5.047	0.0249066
Power of 5	1	42.7	42.7	1.027	0.3112273
Power of 6	1	26.7	26.7	0.641	0.4233990
Power of 7	1	245.7	245.7	5.908	0.0152684
Power of 8	1	52.5	52.5	1.263	0.2613394
Power of 9	1	35.8	35.8	0.861	0.3538391
Power of 10	1	31.1	31.1	0.749	0.3871409
Power of 11	1	4.8	4.8	0.116	0.7339592
Power of 12	1	0.1	0.1	0.003	0.9595070
Power of 13	1	1.8	1.8	0.043	0.8351457
Power of 14	1	0.8	0.8	0.019	0.8895168
Interaction	196	50461.5	257.5	6.190	**0.0000000**
Interaction adjusted level of significance = 0.00005127591. *Only significant results shown.*					
Power of 1: Power of 1	1	34688.8	34688.8	833.966	**0.0000000**
Power of 2: Power of 1	1	1464.2	1464.2	35.200	**0.0000000**
Power of 1: Power of 2	1	5426.3	5426.3	130.457	**0.0000000**
Power of 1: Power of 3	1	925.8	925.8	22.257	**0.0000028**

TABLE XXIX
F6-PARTITIONED SUM OF SQUARES OF POOLED ANALYSIS FOR CROSSOVER

Parameter	Df	Sum of Sq	Mean Sq	F Value	Pr(F)
Crossover	15	818.36	54.56	1.890	0.0207598
Crossover adjusted level of significance = 0.000669798					
Power of 1	1	381.88	381.88	13.232	**0.0002900**
Power of 2	1	7.33	7.33	0.254	0.6143782
Power of 3	1	0.68	0.68	0.024	0.8778748
Power of 4	1	54.75	54.75	1.897	0.1687276
Power of 5	1	37.90	37.90	1.313	0.2520953
Power of 6	1	35.89	35.89	1.243	0.2650954
Power of 7	1	1.05	1.05	0.037	0.8484232
Power of 8	1	23.91	23.91	0.828	0.3629396
Power of 9	1	3.03	3.03	0.105	0.7461390
Power of 10	1	0.10	0.10	0.003	0.9528493
Power of 11	1	18.28	18.28	0.634	0.4262661
Power of 12	1	50.86	50.86	1.762	0.1846610
Power of 13	1	193.18	193.18	6.693	0.0098245
Power of 14	1	4.52	4.52	0.156	0.6925059
Power of 15	1	4.99	4.99	0.173	0.6776497

TABLE XXX
EQUATIONS OF FITTED RESPONSE CURVES

F1	Crossover	Final epoch = $82.35894 - 13.56899\mathrm{Cr}$
	Mutation	Final epoch = $123.5819 - 1830.0797\mathrm{Mu} + 17956.7153\mathrm{Mu}^2 - 43781.1078\mathrm{Mu}^3$
F3	Crossover	Final epoch = $77.99059 - 13.05733\mathrm{Cr}$
	Mutation	Final epoch = $130.9682 - 2707.566\mathrm{Mu} + 26493.42\mathrm{Mu}^2$
F2	Overall	Final epoch = $-1415.7329 + 115.0829\mathrm{Cr} + 30548.5413\mathrm{Mu} - 177255.5477\mathrm{Mu}^2 + 332182.6263\mathrm{Mu}^3 - 428.4953(\mathrm{Cr} * \mathrm{Mu})$
F6	Overall	Final epoch = $163.3295 + 2143.9363\mathrm{Cr} + 222.2216\mathrm{Cr}^2 + 2095.7379\mathrm{Mu} - 30367.4855\mathrm{Mu}^2 + 105193.7584\mathrm{Mu}^3 - 41244.8444(\mathrm{Cr} * \mathrm{Mu}) - 1273.7673(\mathrm{Cr}^2 * \mathrm{Mu}) + 260999.0679(\mathrm{Cr} * \mathrm{Mu}^2) - 543626.2156(\mathrm{Cr} * \mathrm{Mu}^3)$

Crossover parameter level (Cr), Mutation parameter level (Mu).

body of evidence advocating the use of higher mutation rates than have traditionally been used [1]. As with crossover, further statistical work of this kind will assist in the use of the mutation parameter in various problem domains.

The use of statistics also enabled the issue of interaction to be addressed and we found that whether interaction is significant is also problem specific. As to why it is important for some problem domains and not others remains to be answered and may lead to a greater understanding of the interplay between the baseline parameters of crossover and mutation. Again, we are carrying out further research in this area.

In conclusion, this paper has demonstrated a statistical methodology that allows the investigator to undertake exploratory analysis of genetic and other adaptive algorithms. Given the many unique advantages offered by statistical analysis, such as the ability to block for seed, calculation of power and sample size, and rigorous study of response curves, further use of statistics in this exploratory way will assist in the use of GAs as powerful search tools.

APPENDIX A
POWER TABLES

See Tables XVI–XXII.

APPENDIX B
PARTITIONED SUM OF SQUARES

See Tables XXIII–XXIX.

APPENDIX C
FITTED RESPONSE CURVES

See Table XXX.

REFERENCES

[1] T. Bäck, U. Hammel, and H. Schwefel, "Evolutionary computation: Comments on the history and current state," *IEEE Trans. Evol. Comput.*, vol. 1, pp. 3–17, Apr. 1997.

[2] K. A. De Jong, "Analysis of the behavior of a class of genetic adaptive systems," Ph.D. dissertation, Dept. Comput. Commun. Sci., Univ. Michigan, Ann Arbor, MI, 1975.

[3] *Handbook of Genetic Algorithms*, L. Davis, Ed., Van Nostrand, New York, USA, 1991.

[4] J. H. Holland, *Adaptation in Natural and Artificial Systems*. Ann Arbor, MI: Univ. Michigan Press, 1975.

[5] D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*. Reading, MA: Addison-Wesley, 1989.

[6] A. Narayanan and M. Moore, "Quantum inspired genetic algorithms," in *Proc. Int. Conf. Evolutionary Computation*, 1996, pp. 61–66. [Online]. Available: citeseer.nj.nec.com/article/narayanan95quantum inspired.html.

[7] A. B. Simões and E. Costa, "Transposition versus crossover: An empirical study," in *Proc. Genetic Evolutionary Computation Conf. (GECCO)*, W. Banzhaf, J. Daida, A. E. Eiben, M. H. Garzon, V. Honavar, M. Jakiela, and R. E. Smith, Eds., Orlando, FL, July 1999, pp. 612–619.

[8] J. D. Schaffer, R. A. Caruana, L. J. Eshelman, and R. Das, "A study of control parameters affecting online performance of genetic algorithms for function optimization," Philips Lab., North American Philips Corp., Briarcliff Manor, NY, 1989.

[9] L. Eshelman, "Bit-climbers and naive evolution," *Genetic Algorithms Dig.*, vol. 5, no. 39, Dec. 1991.

[10] J. He and X. Yao, "From an individual to a population: An analysis of the first hitting time of population-based evolutionary algorithms," *IEEE Trans. Evol. Comput.*, vol. 6, pp. 495–511, 2002.

[11] T. E. Davis and J. C. Principe, "A markov chain framework for the simple genetic algorithm," *Evol. Comput.*, vol. 1, no. 3, pp. 269–288, 1993.

[12] J. Grefenstette, "Optimization of control parameters for genetic algorithms," *IEEE Trans. Systems, Man, Cybern.*, vol. SMC-16, pp. 122–128, 1986.

[13] B. Freisleben and M. Härtfelder, "Optimization of genetic algorithms by genetic algorithms," in *Proc. 1993 Int. Conf. Aritifical Neural Nets Genetic Algorithms*, Vienna, 1993, pp. 392–399.

[14] X. Yao, Y. Liu, and G. Lin, "Evolutionary programming made faster," *IEEE Trans. Evol. Comput.*, vol. 3, pp. 82–102, July 1999.

[15] A. Petrovski, A. Wilson, and J. McCall, "Statistical analysis of genetic algorithms and inference about optimal factors," School Comput. Math. Sci., Faculty of Sci. Technol., The Robert Gordon University, Aberdeen, U.K., Tech. Rep. 2, SCMS Tech. Rep.1998/2, 1998.

[16] D. H. Wolpert and W. G. Macready, "No free lunch theorems for optimization," *IEEE Trans. Evol. Comput.*, vol. 1, pp. 67–82, Apr. 1997.

[17] L. Davis, "Adapting operator probabilities in genetic algorithms," in *Proc. 3rd Int. Conf. Genetic Algorithms*, D. J. Schaffer, Ed., San Mateo, CA, June 4–7, 1989, pp. 61–69.

[18] R. A. Johnson and G. K. Bhattacharyya, *Statistics: Principles and Methods*. New York: Wiley, 1996.

[19] D. C. Montgomery, *Design and Analysis of Experiments*. New York: Wiley, 1976.

[20] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*. Mahwah, NJ: Lawrence Erlbaum Assoc., 1988.

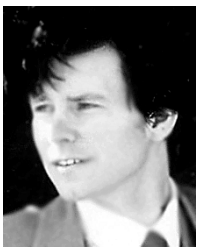
[21] Sun Microsystems, Inc. Java Version 1.3.1_02. [Online]. Available: http://www.sun.java.com/

[22] I. Corporation. S-PLUS. [Online]. Available: http://www.insightful.com/

[23] F. F. and E. E. GPOWER: A priori, post-hoc, and compromise power analyses for MS-DOS [computer program]. Bonn Univ., Dept. Psychology, Bonn, Germany. [Online]. Available: http://www.psycho.uni-duesseldorf.de/aap/projects/gpower/index.html

[24] R. I. Levin and D. S. Rubin, *Applied Elementary Statistics*. Englewood Cliffs, NJ: Prentice-Hall, 1980.

[25] J. A. Rice, *Mathematical Statistics and Data Analysis*. Belmont, CA: Duxbury Press, 1995.



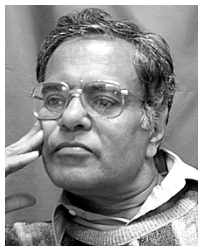
Andrew Czarn received the Ph.D. degree in statistical data mining from the University of Western Australia, Crawley. He is currently working toward the Ph.D. degree in artificial intelligence at the University of Western Australia.

He has held several University, and Public and Private Sector, Research Fellowships and is currently a Researcher with the Adaptive Systems Research Group, School of Computer Science and Software Engineering, University of Western Australia, Crawley. His research interests include interfacing adaptive algorithms and statistical techniques, computational nanotechnology, and computer music.

Cara MacNish received the Bachelor of Engineering degree from the University of Western Australia, Crawley, in 1987 and the Ph.D. degree in logic-based inference systems from the University of Cambridge, Cambridge, U.K., in 1992.

She took a Lectureship in computer science at the University of York, York, U.K. In 1995, she joined the School of Computer Science and Software Engineering, University of Western Australia, Crawley, where she is currently a Senior Lecturer and Convenor of the Adaptive Systems Research Group. Her research interests include logics and artificial intelligence, machine learning, optimization, and computer science education.

Dr. MacNish was awarded the Prince of Wales Studentship to study at Trinity College, Cambridge, U.K., in 1988.



Kaipillil Vijayan received the M.Sc. degree in statistics from the University of Kerala, Kerala, India, in 1961 and the Ph.D. degree in statistics from the Indian Statistical Institute, Calcutta, India, in 1966.

He is a Foundation Fellow with the Institute of Combinatorics and Applications, Winnipeg, Canada. From 1966 to 1968, he was a Lecturer with the Probability and Statistics Department, Sheffield University, Sheffield, U.K. Since 1968, he has been with the University of Western Australia, Crawley, as a Senior Lecturer until 2001, and thereafter, as an Associate Professor. His main research interests are in the areas of sample surveys, (statistical) design of experiments, combinatorics, and graph theory.

Berwin Turlach received the Diplom-Mathematiker degree from the University of Bonn, Bonn, Germany, in 1991, and the Diplome d'Etudes Approfondies en mathematique and Docteur en Statistique degrees from the Universite Catholique de Louvain, Louvain-la-Neuve, Belgium, in 1992 and 1994, respectively.

He is a Lecturer with the School of Mathematics and Statistics, University of Western Australia, Crawley. From January 1995 to June 1998, he was with the Australian National University, Canberra, Australia. The first two years was as a Research Associate with the Centre for Mathematics and its Applications, and then as a Research Fellow with the Cooperative Research Centre for Advanced Computational Systems. From June 1998 to December 1999, he was a Lecturer in the Department of Statistics, University of Adelaide, Adelaide, South Australia. Since December 1999, he has been with the University of Western Australia, Crawley. His research interests include nonparametric smoothing methods and computational statistics.

Ritu Gupta received the M.A. and Ph.D. degrees in statistics from the University of Delhi, Delhi, India, in 1989 and 1994, respectively.

From 1989 to 1996, she was a Lecturer in statistics with the University of Delhi. From 1996 to 1998, she was a part-time Lecturer with the Curtin University of Technology, Australia, and from 1998 to 2000, she was Research Officer with the Statistical Consulting Group, University of Western Australia, Australia. Since 2000, she has been a Lecturer in the Department of Mathematics and Statistics, Curtin University of Technology, Bentley, Australia. Her main research interests are in the areas of combinatorics, queueing systems, and (statistical) design of experiments.